科学推动下技术机会发现方法研究*

■ 韩晓彤　朱东华　汪雪锋
北京理工大学管理与经济学院　北京 100081

摘　要：［目的/意义］科学与技术之间的密切关系，使得结合论文与专利进行技术机会分析比使用单一数据更加合理与高效。本文使科学技术关系的生成更自动化，减弱对主观判断的依赖，使技术颗粒更加精细，同时为技术研发人员提供研发建议，将科学研究中汲取的理念应用到相应的技术创新中去。［方法/过程］将论文、专利的摘要文本进行 Doc2vec 向量表示，通过文本相似性将其关联成网络，再基于 Louvain 算法生成科技主题聚类，识别科学研究推动的技术机会。最后以 3D 打印技术为案例进行实证研究。［结果/结论］识别出若干科学研究推动的技术机会，并验证所识别的机会具有一定的技术潜力，证明方法的可行性和有效性。

关键词：技术机会分析　科学推动力　专利分析　Doc2vec　Louvain 算法
分类号：G250　F204
DOI：10.13266/j.issn.0252－3116.2022.10.002

创新是技术进步与经济发展最重要的源泉之一[1-2]。为合理利用有限的资源，增强创新能力和市场竞争力，科研人员或企业研发人员需要为实现创新而识别各种技术机会。技术机会分析（Technology Opportunity Analysis，TOA）由佐治亚理工学院技术政策与评价中心的 Alan Porter 教授于 20 世纪 90 年代提出，被定义为通过挖掘某领域内已有技术的发展趋势及相互关系，推断该领域即将可能出现的技术形态或技术发展热点。A. L. Porter 教授结合技术监测与文献计量分析，进行了技术机会识别研究[3-4]。在 TOA 被提出后的 20 多年中，全球学者对技术机会分析进行不断的研究，为发现技术机会分析提供了丰富的方法。

现有研究大多基于专利数据进行技术机会分析。专利作为测度产业技术创新的指标，展现了技术发展的成果[5]。在知识密集型技术中，科学研究是技术发展的基础，被认为是技术和创新的种子[6]。科学文献作为科学知识记载、传播与交流的主要方式，反映了科学发展过程和取得的成果[7]。因此，许多学者从科学论文中寻找技术机会。科学与技术的关系密不可分，科学研究为技术创新提供种子，也为技术开发输送养料。从科学推动的视角进行技术机会分析，有利于将科学研究成果应用到技术开发中去。科学研究与技术开发环节相结合的研究通常基于论文与专利两种数据源展开。以往学者们通过混合使用两种数据或分别使用两种数据进行技术机会分析，但都不能较好地利用科学技术之间的关系来挖掘技术机会。

部分学者将论文、专利分别形成科学、技术主题两大聚类并进行对比分析，克服了前两类研究未充分利用科学与技术关系的缺陷。这类研究将论文和专利形成的主题关联起来，将论文中存在而专利中不存在的主题识别为技术机会。一类是分别基于论文、专利引证网络形成科学与技术聚类[6,8-9]。由于发展不成熟的技术还没有形成完善的引证网络，会导致网络稀疏。另一类是基于文本语义分别构建科学聚类、技术聚类，再对比两者主题差异[10-14]。由于两种数据源不同，两种聚类很难对应，对专家的依赖程度大。

为了克服以上局限，本文以科学与技术之间的关系为理论基础，先通过摘要的文本相似性将论文、专利两种数据关联起来，形成关联网络，进而生成科技主题聚类，分析每个聚类中论文和专利的数量及驱动关系，识别科学研究推动的技术机会，从而增强科学、技术主题的对应关系，并减弱了对专家的依赖。

1　相关文献综述

1.1　科学研究与技术开发的交互作用

科学研究与技术开发之间的密切关系，是本文的

* 本文系国家自然科学基金项目“生物医学领域潜在颠覆性技术识别方法研究”（项目编号：72074020）研究成果之一。
作者简介：韩晓彤，博士研究生；朱东华，研究员，博士生导师；汪雪锋，教授，博士生导师，通信作者，E-mail：wxf5122@bit.edu.cn。
收稿日期：2021－11－21　修回日期：2022－01－19　本文起止页码：19－32　本文责任编辑：王传清

理论基础。自1965年M. Price的研究[15]以来,研究人员发现,随着创新周期的缩短,技术与科学研究之间的联系越来越紧密[16-17]。S. Breschi通过科学家与发明家的紧密联系揭示科学与技术的联系[18]。丰富的科学研究可以刺激创新和技术发展[19],从长远来看,新技术机遇最重要的源泉是科学知识的积累。科学可以通过两种方式促进工业技术:一是提供理论、数据和解决问题的能力;二是直接开发新的技术可能性,为老问题提出新的解决方案[20]。如果没有最近的学术研究,就会少10%的新产品和新工艺[21]。相反,科学探索较少的技术也可能激发重要的科学突破,大多数应用研发工作都是从需求或目标开始的,然后再回到科学来实现目标[20]。因此,科学和技术是相互依存的[22-23]。在实证研究方面,赖院根基于主题词对应,将CLC分类号与IPC的类目形成映射,连接论文和专利以探索科学研究与技术研发的关系[24]。许多实证研究表明,大量的科学信息在专利中被组织和编纂,这表明越来越多的技术发展依赖于科学[25-27]。科技期刊论文与专利文献之间的相互引证现象,也表明科学与技术之间、基础研究与技术创新之间相互联系、相互影响与相互促进的关系[28-30]。L. Fleming通过基于专利数据的实证分析,证实科学改变了发明家的寻找过程,使他们消除无效研究路径,更直接地找到有用的组合,并激励他们在面对负面反馈的情况下继续前进[31]。由于科学家们很少知道他们的科学发现会有什么样的工业应用,而公司往往不知道什么科学发现可能对他们的需求有帮助,T. Hellmann构建了科学家和企业之间的搜索和匹配模型,弥合了科学家和公司进行搜索和交流时的障碍,其研究发现,为科学发现申请专利有助于科学界将他们的发现推向工业领域[32]。

1.2 基于科学论文与专利的机会分析研究

基于论文与专利的技术机会分析研究大致可分为3类:

(1)研究不区分论文、专利的数据来源,混合使用。如任智军通过挖掘论文和专利文本,实现了新兴技术发现方法,根据特征选择结果预测技术趋势和新兴技术[33]。

(2)区别分析论文与专利,分别识别两者中的技术机会。如张福俊基于学术论文、专利文献,采用共词分析法和谓词树分析法在论文与专利之间以及专利之间进行对比分析,挖掘出海洋科学领域的技术机会[34]。王兴旺将科技论文、专利和科技舆情这三种不同类型的信息结合起来,识别高暴发强度关键词用于前沿技术预测[35]。

(3)将论文、专利分别形成科学、技术主题,进行对比分析,克服了前两类研究未充分利用科学与技术关系的缺陷。这类研究将论文和专利形成的主题关联起来,将论文中存在而专利中不存在的主题识别为技术机会。一类是分别基于论文、专利引证网络形成科学与技术聚类,如N. Shibata等构建论文、专利引证网络,通过自然语言处理测量了学术论文集和专利集之间的语义相似性,挖掘只存在于学术研究而没有专利的前沿领域,以支持企业与政府的研发投资决策[6,8]。Y. Takano和Y. Kajikawa分别构建论文、专利引证网络,计算论文聚类、专利聚类之间的余弦相似性,选出两者相似度低、论文多而无专利的新兴聚类,进行技术机会分析[9]。另一类是基于文本语义分别构建科学聚类、技术聚类,再对比两者主题差异。如M. Y. Wang通过挖掘论文和专利,形成科学、技术领域聚类,识别出有科学活动但没有技术应用的领域,提供了新技术机会的潜力[11]。黄鲁成、王静静抽取论文和专利文本中的SAO结构,计算论文相似度和专利相似度,通过多维尺度分析总结聚类主题,识别论文中存在而专利中不存在的主题以确定技术机会[13]。王坤等提取论文、专利文本关键词并构建相异共现矩阵,采用多维聚类方法识别研究热点主题,将论文中存在而专利中不存在的主题识别为技术机会[14]。王京安对比分析科技论文和专利文献中关键词聚类、技术研究热点,发现物联网发展的两大类技术机会分别是在物联网应用层上GIS技术和平台层中多源大数据下融合计算技术[12]。韩燕运用连边关键程度中心度指标、知识单元的整合等方法提取技术主题,考虑科学研究与技术成果这两种技术形成的基础对技术主题进行分析,最终找到以"知识关联"为表现形式的技术机会[36]。X. Li通过比较技术主题首次出现在科学论文和专利中的时间差异,结合文本挖掘和专家判断,预测技术短期发展趋势,识别技术机会[10]。整理归纳第三类研究见表1。

1.3 文献评述

技术机会分析的大多数研究都使用单一的数据(科学论文或专利),在分析过程中缺失了许多有用的技术信息,使分析结果不完善。科学研究与技术开发之间存在着密切的关系,科学研究对技术研发具有推动力。若将科学论文与技术专利结合,较单独使用科学论文或专利,有利于更全面地进行技术机会分析。然而许多研究在进行论文和专利的联合分析时,尽管同时使用了两种数据、涵盖了更多科技信息,但仍旧是

表 1　基于科学论文与专利关系的技术机会分析研究

作者	主题形成方式	主题关联方式	研究内容
N. Shibata, et al. (2010,2011)	引证网络拓扑聚类	人工判断 语义相似性	挖掘只存在于学术研究而没有专利的前沿领域
Y. Kajikawa & Y. Kajikawa(2019)	专利家族引证网络递归聚类	余弦相似性	选出相似度低、论文多而专利少的新兴聚类
M. Y. Wang, et al. (2015)	ORCLUS 聚类算法	人工判断	识别有科学活动但无技术应用的领域
黄鲁成等(2016)	SAO 相似度聚类	人工判断	识别论文存在而专利不存在的主题
王京安等(2018)	Citespace 共词网络聚类	人工判断	识别没有科学论文的技术热点、没有专利的科学热点
王坤等(2018)	关键词多维聚类	人工判断	没有科学论文的技术主题、没有技术专利的科学主题
韩燕和彭爱东(2020)	知识单元整合	人工判断	找到能对应的科学主题与技术主题， 即以“知识关联”为表现形式的技术机会
X. Li, et al. (2019)	Lingo 聚类算法	时间差异人工判断	比较技术主题首次出现在科学论文和专利中的时间差异， 预测技术短期发展趋势，识别技术机会

分别对两种数据进行分析，将其割裂开来，缺少两者之间的对比与联系，也并未将两者的关系考虑进去。

近年来，一些学者在技术机会分析的研究(见表1)中，通过分别聚类论文和专利，得到科学主题和技术主题，并对比分析，克服了以往研究割裂两种数据的缺陷。然而，上述研究也存在局限性：①在基于引证网络形成科学、技术主题的研究中，由于刚兴起的技术还没有形成完善的引证网络，在网络构建时会被排除，导致网络稀疏，也可能会丢失重要信息；②论文与其参考文献不一定讲述同一主题。在基于文本形成主题聚类的研究中，由于两种数据聚类所使用的文本不同，两种聚类的对应程度比较低，需要靠专家进行聚类的对应识别，最终的结果技术粒度较粗，且通过这种方法所能匹配上的科学－技术主题对往往只有寥寥数个，不利于挖掘更多的技术机会。

2　基于科学推动力的技术机会分析方法

本文研究框架如图 1 所示，分为 3 个步骤：①数据收集与文档向量表示：选择合适技术领域，收集科学论文与专利数据，提取摘要文本，基于 Doc2vec 模型将论文、专利表示成“文档－特征向量”的形式；②科技主题网络的构建：基于“文档－特征向量”计算各文档之间的余弦相似性，从而构建科学论文与专利之间的关系网络，通过 Louvain 社区发现算法形成科技主题聚类，将科学论文－专利网络分割成若干子网络；③技术机会发现：构建“相对科学推动力”与“相对技术成果数量”指标，建立二维坐标系，识别潜在的技术机会。

科学论文数据来自 SCI(科学引文)索引与 EI(工程)索引。SCI 是国际公认的进行科学统计与科学评价检索工具，EI 则是著名的工程技术类综合性检索工具，两者收录的科学论文有差异也有交叉，但都是本文研究所需的表征科学过程和成果的文档。本文以科学论文标题为匹配依据，将 SCI 与 EI 检索所获取的文档去除重复记录，即取两者的并集作为研究所用的专利数据集。技术专利数据来自专利数据平台德温特创新索引(Derwent Innovation Index，DII)。提取论文与专利数据的标题、摘要文本与年份信息，作为本文分析之用。

为了使论文与专利数据关联匹配，需要用文档向量表示方法将两者摘要文本映射到同一向量空间。过去常用的文档向量表示方法有以下几种：①Bag of Words，即词袋模型；②Average Word Vectors，对句子中的所有词向量简单地取平均；③TF-IDF Weighted Word Vectors，对句子中的所有词向量根据 TF-IDF 权重加权求和。这 3 种方法共同的缺点是都不考虑单词的顺序，忽略了语义信息。Doc2vec[37] 是目前较有效的文档向量表示方法，它是一种无监督算法，基于单词的上下文顺序从文档中学习得到固定长度的特征表示，用于预测一个向量来表示不同的文档，克服了 Bag of Words 等模型忽略单词顺序的缺点。Doc2vec 模型是受 Word2Vec 词向量模型启发而来，与 Word2vec 相似，Doc2vec 也有两种训练方式——PV-DM(Distributed Memory Model of Paragraph Vectors)和 PV-DBOW(Distributed Bag of Words version of Paragraph Vector)。

2.1　数据收集与 Doc2vec 文档向量表示

本文使用 python 的 gensim 库进行文档向量训练。根据模型作者的推荐[37]，以相同的参数分别用两种模型训练得到文档向量，再将两组向量组合使用。为了使论文向量与专利向量处于同一空间，将两种数据源在标记后放在一起训练。在训练之前，为了得到更好的效果，将摘要文本中同时出现频率较高的两个邻近词组合为一个二元组，作为一个单词进行模型的训练。

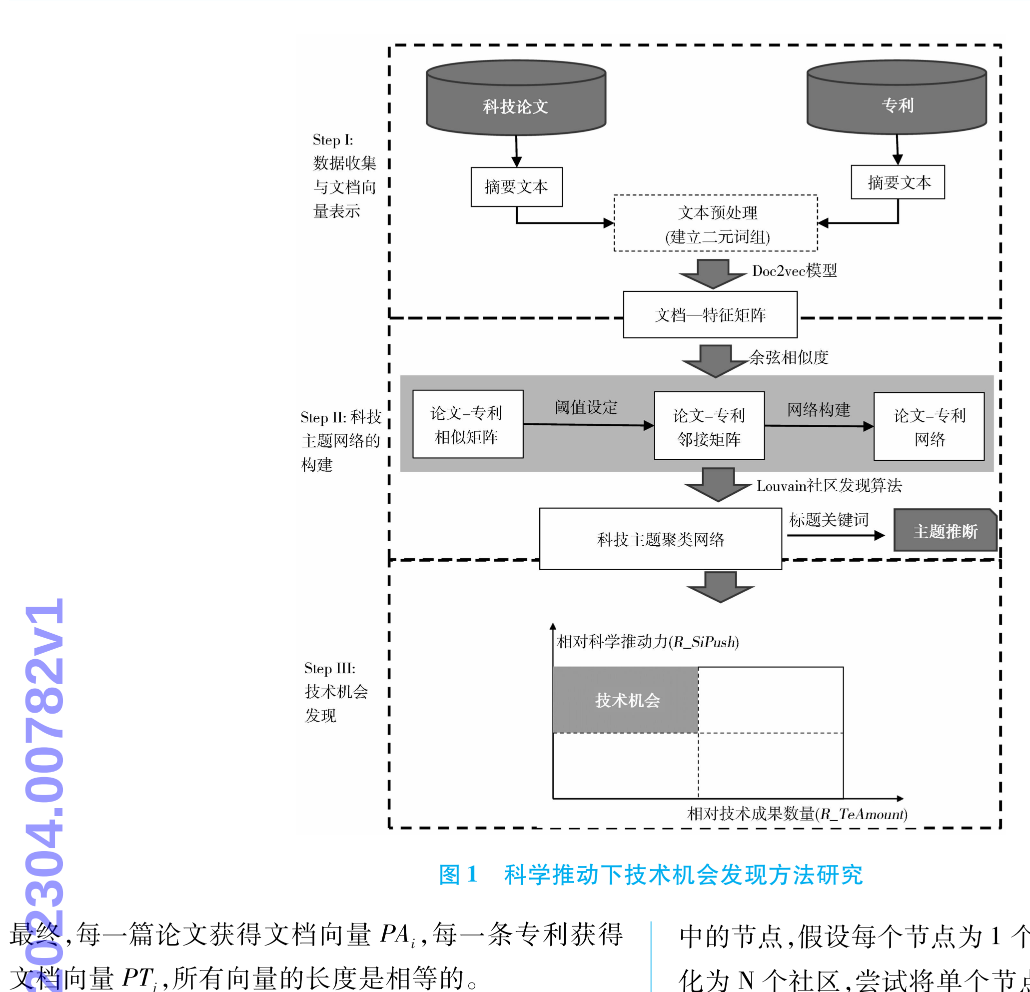


图 1 科学推动下技术机会发现方法研究

最终,每一篇论文获得文档向量 PA_i,每一条专利获得文档向量 PT_j,所有向量的长度是相等的。

2.2 基于 Louvain 社区发现算法的科技主题网络构建

将论文与专利表达为文档向量后,计算每一条论文与专利文档之余弦相似度:

$$cossim(i,j) = \frac{PA_i \cdot PT_j}{|PA_i||PT_j|} \qquad 公式(1)$$

余弦相似度的值分布在 0 – 1 之间,$cossim(i,j)$ 的值越接近 1,说明论文 i 与专利 j 在内容上越相似,越接近 0,说明两者越无关。为了提取论文与专利之间的关联,设定一个阈值,排除余弦相似度小于阈值的关系对,而将余弦相似度大于等于该阈值的论文 – 专利对作为论文 – 专利关联网络中的边,重新设定边权重为 1,从而构建论文 – 专利关联网络。

论文 – 专利关联网络错综庞杂,不利于解读,需将其聚类成可以解读的主题,由于网络中既有代表科学的论文节点,又有代表技术的专利节点,本文将其聚类后的主题称为科技主题。

Louvain 社区发现算法[38]是社会网络中常用的聚类方法,它实际上是基于层次聚类的一种图聚类方法,包括交替迭代的两个阶段。第一阶段不断地遍历网络中的节点,假设每个节点为 1 个社区,N 个节点则初始化为 N 个社区,尝试将单个节点加入能够使模块度提升最大的社区中,直到所有节点都不再变化。模块度是衡量社区分区质量的指标,值越高,则社区内部的连边多、社区外部的连边少,即更聚集。在加权网络的情况下,模块度的计算公式如下:

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \qquad 公式(2)$$

其中 A_{ij} 表示 i 和 j 之间边的权重,$k_i = \sum_j A_{ij}$ 是连接节点 i 的边的权重之和,c_i 是顶点 i 被分配到的社区,$\delta(c_i, c_j)$ 是一个自定义函数,如果 $c_i = c_j$,则 $\delta(c_i, c_j)$ 取值为 1,否则为 0,且 $m = \frac{1}{2}\sum_{i,j} A_{ij}$。

在第二阶段中,将一个个小的社区归并为一个新的节点来重新构造网络。迭代这两个步骤直至算法稳定,即所有子社区模块度总和不变。

在实际研究过程中,进行一次 Louvain 算法后,存在少数几个规模极大的聚类,且涵盖了网络中绝大部分的节点。为了使主题粒度更小更精确,分别对规模很大的聚类进行二次计算,同样使用 Louvain 算法,而其他小类仍使用第一次聚类的结果,最终形成数量合适的科技主题聚类。每一个科技主题都以网络的形式

存在,其中包含论文节点与专利节点,是论文－专利关联网络的子网络。

2.3 基于科学推动力的技术机会发现

在本文中,将科学推动力(*SiPush*)定义为:在某个科技主题聚类网络中,存在一对相连的论文－专利对,专利的发表年份晚于论文至少 n 年,这对关系则称为 1 个科学推动力,推动力的个数即为该主题中 *SiPush* 的值。

M. P. Carpenter 等在学术论文和专利的关联研究中,分析了专利引用学术论文的时滞,专利引用论文通常在论文发表后的 3－5 年(本文称之为引用时滞),与科学论文之间的引用相似[39]。X. Li 从专利和论文中分别提取技术主题和科学主题,研究发现同一主题在专利中首次出现比在论文中首次出现晚 1－2 年(本文称之为主题时滞)[10]。在已有研究中,主题时滞与引用时滞差别较大,因为主题时滞基于内容的相似性,引用时滞基于引用关系,前者表示科学、技术在相对独立发展的情况下的时差,后者却直接体现了科学对技术推动作用的时滞,即科学研究的成果在技术研发中体现至少需要 3 年时间。因此在论文与专利关联中,只有当专利申请年份晚于论文发表时间 3 年及以上,才能说明该技术发明的推动力来自科学研究,而不是相对独立的发展。本文研究的是科学推动的技术机会,因而将技术机会的科学推动力来源限定在早于专利申请时间 3 年及以上(即 n＝3)的科学论文中。

假设一个主题中的论文数量为 *SiNum*,构建以下两个指标:

(1)相对科学推动力(*R_SiPush*),表示一个科技主题中,科学推动力的相对大小,计算公式为:

$$R_SiPush = \frac{SiPush}{SiNum} \qquad 公式(3)$$

其中,*SiPush* 为该主题中科学推动力的数量。

(2)相对技术成果数量(*R_TeNum*),表示科技聚类中,现有技术成果的相对多少,计算公式为:

$$R_TeNum = \frac{TeNum}{SiNum} \qquad 公式(4)$$

其中,*TeNum* 为该主题中专利的数量。

如图 2 所示,图中网络代表一个科技主题聚类,空心节点代表论文,数量为 7,实心节点代表专利,数量为 6,科学推动力(即专利发表年份晚于论文 3 年及以上的连边)个数为 12,则 *R_SiPush*＝12/7＝1.71,*R_TeNum*＝6/7＝0.86。

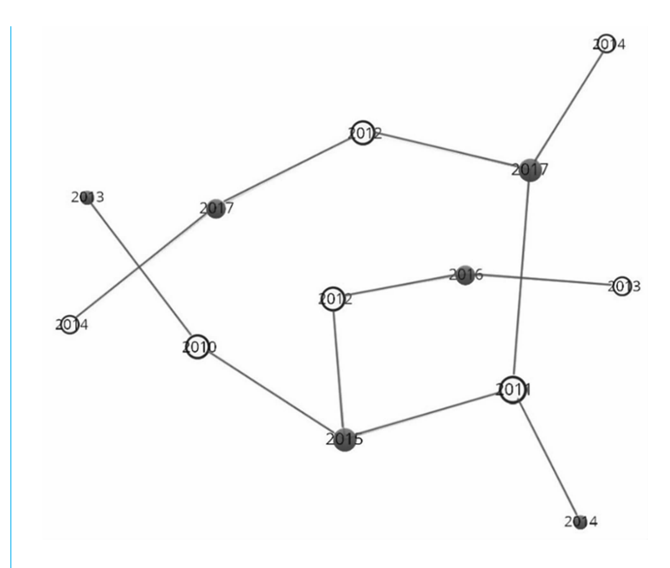


图 2 相对科学推动力、相对技术成果的数量计算示例

用上述两个指标建立二维象限,如图 3 所示。横坐标代表相对技术成果数量,科技主题位于越右端位置,表示在科学论文数量相同的情况下,该主题的专利数量越多,技术实现度越高;纵坐标代表相对科学推动力,科技主题位于越上端的位置,表示在相同科学论文数量下,该主题中专利与论文时差在 3 年及以上的关联数量越多,说明科学论文的推动力相对较大。

第一象限(右上角区域),科学推动力与现有技术成果都相对较多。该区域中的主题,技术的实现对科学研究依赖程度较大,且技术实现程度较高,技术发展相对趋于成熟,未来可开发的技术机会可能相对较少,不适合作为技术机会挖掘的区域。

第二象限(左上角区域),科学推动力大,但现有技术成果少。该区域中的主题,技术实现对科学研究的依赖程度大,但已实现的技术成果并不多,未来技术被开发的可能性较大,且易于从科学研究中汲取技术开发需要的理论基础。该区域可用于挖掘基于科学推动力的技术机会。

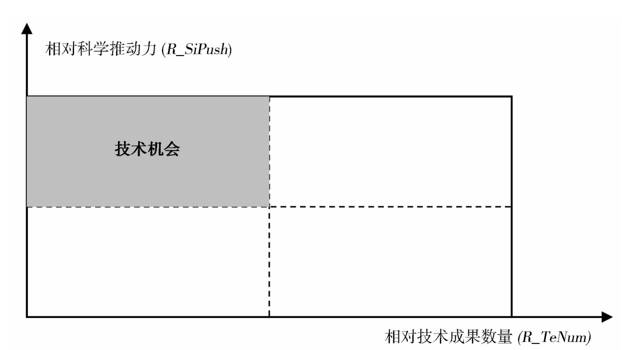


图 3 技术机会二维象限图

第三象限(左下角区域),科学推动力与现有技术成果都相对较少,技术实现对科学研究的依赖程度、技术本身的实现度都较低,可能是科技主题处于起步阶段或衰落阶段,可从科学研究中汲取的理论与理念也相对较少。该区域的主题不在本文讨论范围内。

第四象限(右下角区域),科学推动力较少,现有技术成果较多。表明技术已趋于成熟,未来可发展的技术机会相对较少。较少的科学推动力表明该主题中技术实现对科学研究的依赖程度较低,或技术已发展成熟、无需大量的科学研究支持。该区域也不在本文讨论范围之内。

基于本文的研究目标,选取第二象限作为发现技术机会的区域。

3 实证研究——3D 打印领域技术机会分析

3D 打印(three-dimensional printing),又称增材制造(additive manufacturing),是一种快速成型技术,是当前发展迅速的热门技术领域之一。3D 打印技术经过数十年的发展,已有一定基础,且当前还处于快速发展的阶段,科学论文与技术专利都有可观的数量并持续上升。在商业领域,应用广泛,是本研究较为合适的案例。

3.1 数据检索及文档向量模型

以 2010-2019 年为检索区间,参考以往学者的相关成果[40],拟定检索式"TI = (((3D OR 3-D OR (3 ADJ2 dimension*) OR (three ADJ2 dimension*) OR additive) ADJ (print* OR fabricat* OR manufactur*)))",分别使用 SCI、EI、Derwent 索引进行论文、专利数据的检索,检索时间为 2021 年 2 月 9 日。SCI、EI 论文合并、去重、删去无效数据后共 23 018 篇,德温特专利删去无效数据后共 22 250 条。根据本文研究设计,专利申请年份晚于论文 3 年及以上才视为科学对技术具有推动力,另预留 2018-2019 年数据作为结果验证,所以本文中论文数据涉及年份范围为 2010-2014 年,专利为 2013-2017 年。

经过反复尝试,并结合模型作者设定的默认参数[37],设定 Doc2vec 模型文档向量维数为 100、滑动窗口长度为 10、单词最小频率为 4,并将文本中相邻出现频率大于或等于 20 次的两个单词组合成二元组。

3.2 科技主题网络构建

将论文、专利表示成文档向量之后,计算论文、专利两两之间的余弦相似度,取阈值为 0.5,相似度大于等于阈值的论文-专利对视为有内容关联,记为一条连边,由此构建论文-专利关联网络。

基于论文-专利关联网络,用 Louvain 社区发现算法进行聚类。第一次聚类后,共发现 5 个较大的社区,分别包含 11 395、4 635、7 634、10 689、5 769 个节点,其余社区节点数均不足 60 个。将上述 5 个类团进行二次聚类,排除不具有科学推动力的社区,共得到有效科技主题 107 个。

3.3 技术机会发现

计算 107 个科技主题聚类的相对科学推动力与相对技术成果数量,取两者的中位数 $median_{R_SiPush}$ = 0.410、$median_{R_TeNum}$ = 0.960 作为划分象限的阈值,由于存在极值点"1_14",对其他点的可视化展示造成影响,所以将该点排除,绘制如图 4 所示的二维散点图,图中一个点代表一个科技主题,标签为其聚类编号,"-"前的数值代表第一次聚类的结果,"-"后的数值代表第二次聚类的结果。

处于第二象限的科技主题共 23 个,科学推动力相对较大,现有的技术成果相对较小,是本文识别技术机会的目标区域。这 23 个科技主题中科学推动力(*SiPush*)、论文数量(*SiNum*)、专利数量(*TeNum*)、相对科学推动力(*R_SiPush*)、相对技术成果数量(*R_TeNum*)的统计结果见表 2。

选取规模较大的 10 个科技主题为示例进行结果解读与验证,由于主题"5_9""3_9""5_10"意义较为模糊,故没有选取。根据论文、专利标题中的关键词来推断主题内容,各主题的高频词(含科学词、技术词两类,分别来自论文、专利的标题)如表 3 所示,限于篇幅,仅展示频率最高的 5 个词。排序方式采用加权频率的方法,即标题词的频率以某个词在论文-专利关联对中出现的次数计算,如文档重复出现,标题词也会重复计算频率,因为存在关联较多的文档,理论上在聚类中扮演更重要的角色,因此其标题词的权重也重复计算。科技主题内容推断结果见表 3,即为本文所识别的技术机会。各主题的高频科学词、技术词可视化见图 5,科学词以浅灰色的节点表示、节点标签以"[si]"开头,技术词以深灰色节点表示、节点标签以"[te]"开头。

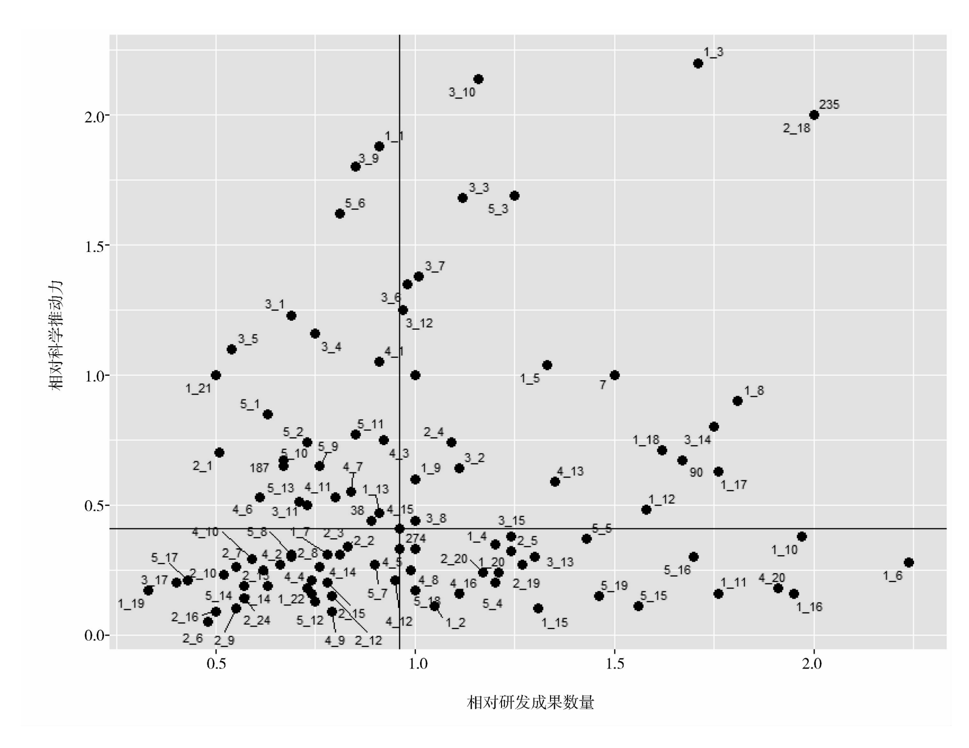


图4　技术机会识别二维象限图

表2　第二象限聚类相关数据统计

主题	*SiPush*	*SiNum*	*TeNum*	*R_SiPush*	*R_TeNum*
1_1	1 024	544	496	1. 88	0. 91
5_6	320	197	159	1. 62	0. 81
3_1	313	255	175	1. 23	0. 69
4_1	308	293	268	1. 05	0. 91
3_9	302	168	143	1. 8	0. 85
3_4	284	244	184	1. 16	0. 75
2_1	210	302	153	0. 7	0. 51
4_7	201	368	309	0. 55	0. 84
5_10	148	229	153	0. 65	0. 67
5_9	131	201	152	0. 65	0. 76
4_6	125	238	145	0. 53	0. 61
5_13	124	241	172	0. 51	0. 71
3_5	113	103	56	1. 1	0. 54
5_11	100	130	111	0. 77	0. 85
5_1	96	113	71	0. 85	0. 63
5_2	96	130	95	0. 74	0. 73
4_11	93	176	141	0. 53	0. 8
3_11	33	66	48	0. 5	0. 73
1_13	15	32	29	0. 47	0. 91
4_3	9	12	11	0. 75	0. 92
38	4	9	8	0. 44	0. 89
187	2	3	2	0. 67	0. 67
1_21	2	2	1	1	0. 5

表 3　科技主题聚类高频词表与主题内容推断

聚类	论文高频标题词		加权频率	专利高频标题词		加权频率
1_1　3D 打印显示设备	photonic crystal	光子晶体	609	memory device	存储设备	628
	holographic fabrication	全息制作	378	circuit board	电路板	392
	3d photonic crystal	三维光子晶体	361	liquid crystal	液晶	357
	lens antenna	透镜天线	326	display device	显示设备	321
	millimeter wave	毫米波	311	semiconductor device	半导体器件	308
3_1　3D 打印电极	3d graphene	三维石墨烯	448	active material	活性物质	754
	graphene oxide	石墨烯氧化物	404	organic solvent	有机溶剂	531
	performance supercapacitor	超级电容器性能	403	lithium salt	锂盐	455
	carbon nanotube	碳纳米管	371	lithium ion battery	锂离子电池	453
	reduced graphene oxide	降低石墨烯氧化物	365	functional additive	功能性添加剂	388
4_1　陶瓷 3D 打印	binder jet	粘结剂喷射	449	sintering additive	烧结添加剂	903
	binder jetting	粘结剂喷射	396	silicon nitride	氮化硅	711
	laser additive manufacturing	激光增材制造	277	pressureless sintering	无压烧结	386
	alumina ceramic	氧化铝陶瓷	210	ceramic material	陶瓷材料	370
	particle size distribution	粒度分布	195	lanthanum oxide	氧化镧	303
	powder packing	粉末包装	195	silicon carbide	碳化硅	284
3_4　3D 喷墨打印	inkjet printing	喷墨打印	163	epoxy resin	环氧树脂	433
	current collector	集电器	130	printing ink	印刷油墨	433
	patterned 3d silicon	有图案的三维硅	129	circuit board	电路板	388
	patch antenna	贴片天线	120	conductive ink	导电油墨	347
	aerosol jet	气溶胶喷射	115	printed circuit	印刷电路	346
2_1　骨组织 3D 打印	calcium phosphate	磷酸氢钙	541	bone tissue	骨组织	383
	bone tissue	骨组织	474	3d printing base	3d 打印技术基础	225
	tricalcium phosphate	磷酸三钙	422	bone defect	骨缺损	225
	bone regeneration	骨再生	420	calcium phosphate	磷酸氢钙	224
	bone tissue engineering	骨组织工程	364	bone repair	骨修复	220
4_7　粉床 3D 打印	powder bed	粉床	519	powder bed	粉床	203
	powder bed fusion	粉床融合	417	operational flaw	操作缺陷	141
	powder bed fusion additive	粉床熔合增材	372	coordinate position	坐标位置	132
	laser powder	激光粉末	323	laser additive	激光增材	125
	metal additive	金属增材	256	laser additive manufacturing	激光增材制造	124
5_9　食物 3D 打印	food additive	食品添加剂	97	feed additive	饲料添加剂	297
	extrusion 3d printing	挤压 3d 打印	71	food additive	食品添加剂	279
	3d food	3d 食物	64	fresh lemon juice	新鲜的柠檬汁	191
	3d food printer	3d 食物打印机	49	momordica grosvenori extract	罗汉果提取物	191
	food printer	食物打印机	49	rich nutrition	营养丰富	191
4_6　激光金属 3D 打印	laser additive manufacturing	激光增材制造	478	alloy powder	合金粉末	882
	powder bed	粉床	259	aluminum alloy	铝合金	424
	powder bed fusion	粉床融合	254	titanium alloy	钛合金	242
	selective laser	选择性激光	215	high temperature	高温	206
	laser melting	激光熔化	191	powder composition	粉成分	180
5_13　3D 打印设备质量控制、3D 打印云服务系统	fault diagnosis	故障诊断	75	3d model	三维模型	184
	delta 3d printer	δ3 d 打印机	64	control module	控制模块	147
	additive manufacturing security	增材制造安全	58	cloud server	云服务器	101
	nanocellulose ink	纳米纤维素墨水	58	control system	控制系统	94
	security awareness perspective	安全意识的角度	58	inventory spare part	库存备件	92
3_5　3D 打印复合材料	polymer composite	聚合物复合材料	72	composite material	复合材料	361
	carbon fiber	碳纤维	59	lactic acid	乳酸	279
	lactic acid	乳酸	54	polylactic acid	聚乳酸	279
	surface layer	表层	51	toughening agent	增韧剂	235
	bone regeneration	骨再生	47	inorganic filler	无机填料	218

韩晓彤，朱东华，汪雪锋. 科学推动下技术机会发现方法研究[J]. 图书情报工作,2022,66(10):19－32.

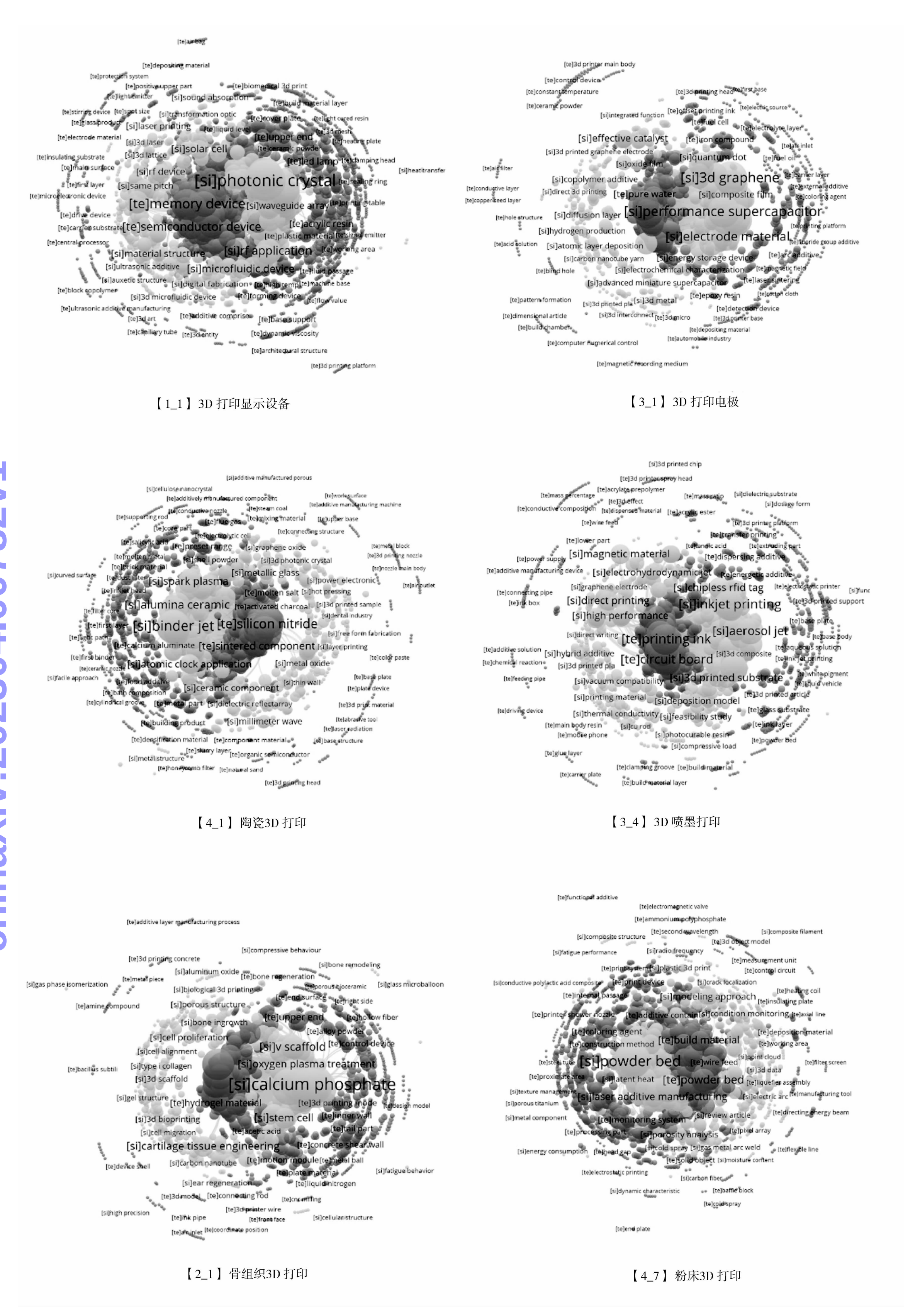


图 5(A)　技术机会主题词网络

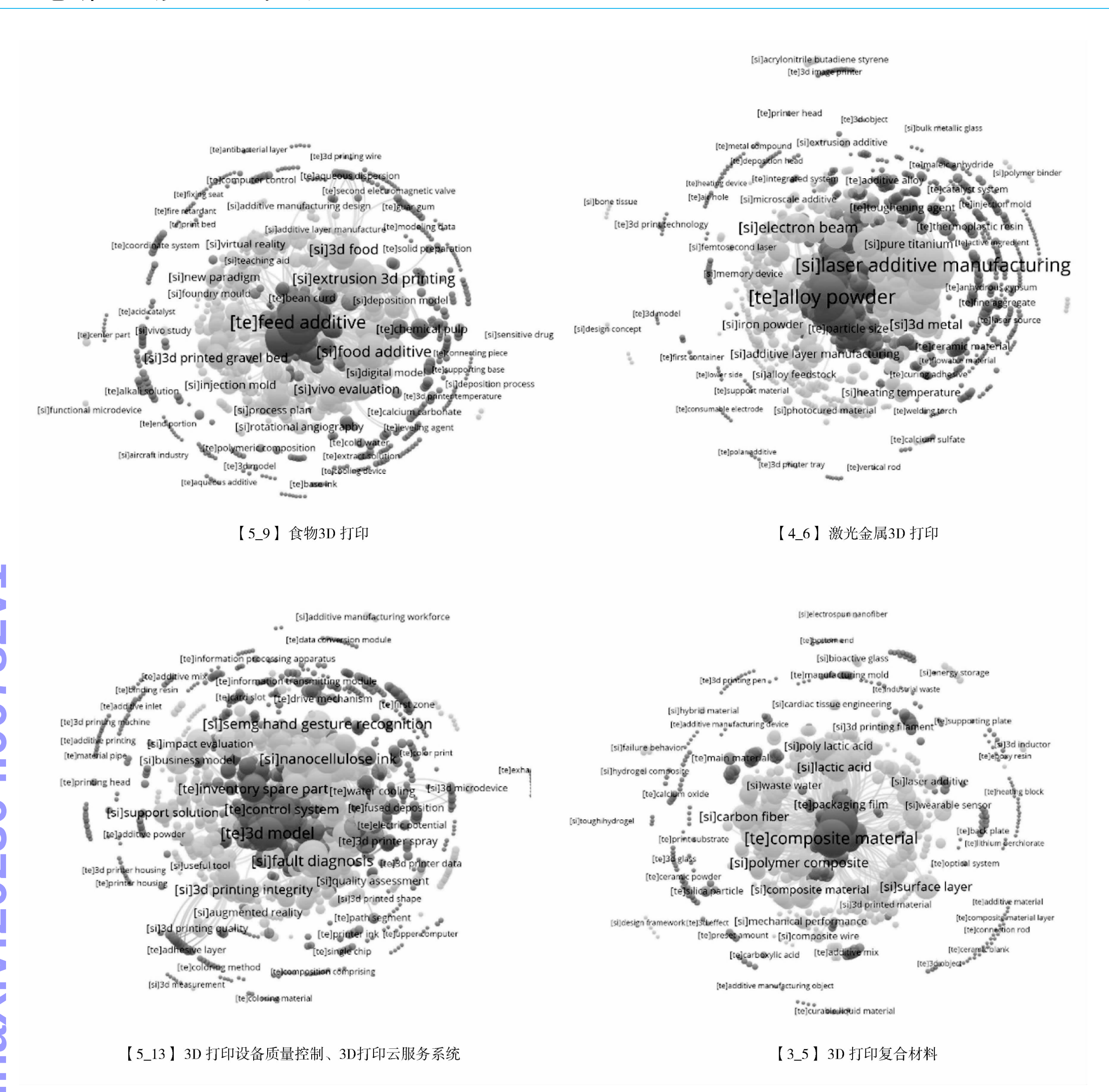


图 5(B) 技术机会主题词网络(续)

主题“1_1”为 3D 打印显示设备，根据“全息制作”“三维光子晶体”等高频论文标题词与“液晶”“显示设备”等高频专利词，技术开发人员可从科学研究中挖掘全息技术的理论基础，开发 3D 打印全息显示屏。主题“3_1”为 3D 打印电极，根据“超级电容器性能”“高性能”“电化学性能”等论文标题词语与“有机溶剂”“电极材料”等专利标题词，技术机会存在于各种有机溶剂、电极材料，科学研究的推动，有望使其向高性能方向发展。主题“4_1”为陶瓷 3D 打印，技术人员可借鉴科学研究，从陶瓷材料、烧结工艺等方面对技术进行创新或提高。主题“3_4”是 3D 喷墨打印，技术人员可在印刷油墨等方向进行技术开发，并从科学研究中获取相关知识。主题“2_1”是骨组织 3D 打印，技术人员可从科学研究追踪汲取知识与理念，开发或改进骨组织 3D 打印技术，从骨再生、骨组织修复等技术中找到技术机会。主题“4_7”是粉床 3D 打印，技术人员可从粉床、激光 3D 打印技术等方面的技术机会中进行技术创新。主题“5_9”是食物 3D 打印，科学研究的发展推动技术的进步，从食品打印原材料提取、食品与饲料 3D 打印技术等方面带来技术机会。主题“4_6”是激光金属 3D 打印，金属粉末材料、激光 3D 打印技术、记忆材料等方面存在技术机会，同时可能科学研究成果中借鉴相关理念。主题“5_13”是 3D 打印设备质量控制、3D 打印云服务系统，从科学研究中掘取故障诊断理念、安全知识，应用到提升 3D 打印设备的质量与完善 3D 打印云服务系统中去。主题“3_5”是 3D 打印复合材料，可借鉴科学研究中的各种化学实验、材料研究，进行 3D 打印复合材料的创新与改进。

3.4 结果验证

为验证本方法所识别的机会是否具有较大的技术

潜力，以及技术潜力是否在较大程度上由科学研究推动，将各技术主题 2013－2019 年的专利数量、以及具有科学推动力的专利数量进行统计展示，如图 6 所示，其中直方图代表该科技主题中的专利数量，折线图代表该主题中具有科学推动力的专利数量。图 6 中所有主题的专利趋势呈快速上升趋势（直方图），科学推动的趋势也快速增长（折线图），个别主题的专利数量在 2019 年看似有所回落，可能是因为专利申请到公开具有一定的时间延迟，导致德温特数据库收录的 2019 年专利数据不完整。此外，具有科学推动力的专利数（折线图）占专利总数（直方图）的比重超过大半，且该比例有上开趋势。总体来说，该统计结果可以说明本文方法所识别的技术机会确实具有技术潜力，且在较大程度上由科学研究推动，验证了方法的有效性。

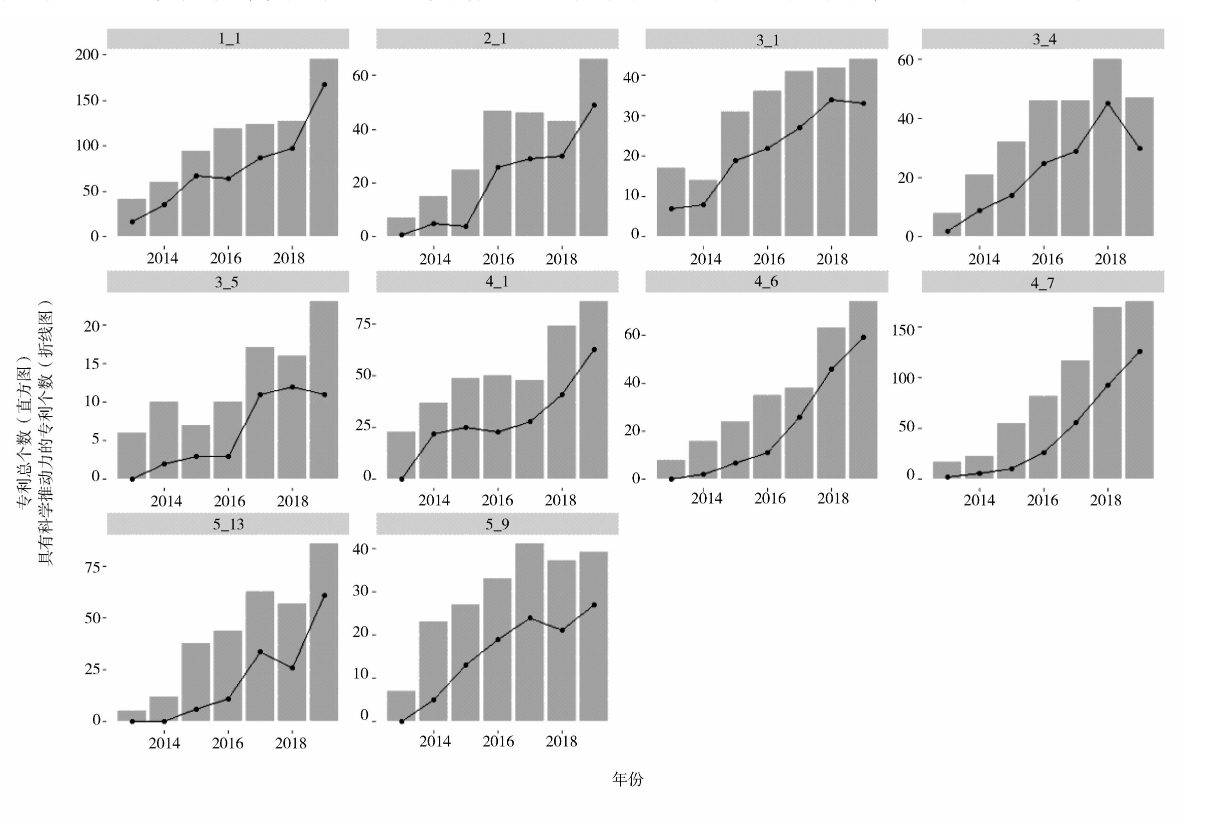


图 6　各技术机会 2013－2019 年专利数量走势

各技术机会在科技主题散点图中处于第二象限。为了对比该象限与其他象限的专利增长趋势，绘制了第一、三、四象限的专利数与具有科学推动力的专利数趋势图（见图 8）。在第一象限中，尽管科学推动力（折线图）的占比较高，但专利总数（直方图）不具有明显的上升趋势，该区域的科技主题可能处于较成熟的发展阶段，产生技术机会的可能性不大；第三、四象限中，具有科学推动力的专利数（折线图）占专利总数（直方图）的比重始终较低，说明这两个象限中的技术发展受科学研究的影响相对较小，且第四象限的专利总数不具有明显上升趋势，表明这两个象限中的技术在科学推动下的发展潜力不大。

在其他学者的 3D 打印领域技术机会分析研究中，王金凤[41]通过专利文本挖掘和形态分析，识别出切换式多色打印机头、激光烧结、复合材料的组合为最佳技术机会。王京安[14]从科学论文与专利相结合的角度，识别出科学理论专利化、产业化的机会：用直接激光金属烧结和沉积技术改变材料特性、通过非线性建模仿真测试金属材料特性。以上结果与本文识别的激光金属 3D 打印、3D 打印复合材料、陶瓷 3D 打印等主题或主题中包含的技术元素部分吻合，且本文识别的技术机会范围更广、数量更多，从另一个角度验证了本文方法的效果和优势。

4　总结与讨论

本文从科学研究与技术开发环节结合的角度，介绍了科学推动下的技术机会发现方法。该方法以科学论文、技术专利为数据源，用 Doc2vec 模型将两者摘要文本表示成同一空间中的向量，通过计算余弦相似度将两者关联，形成论文－专利关联网络，进而用 Louvain

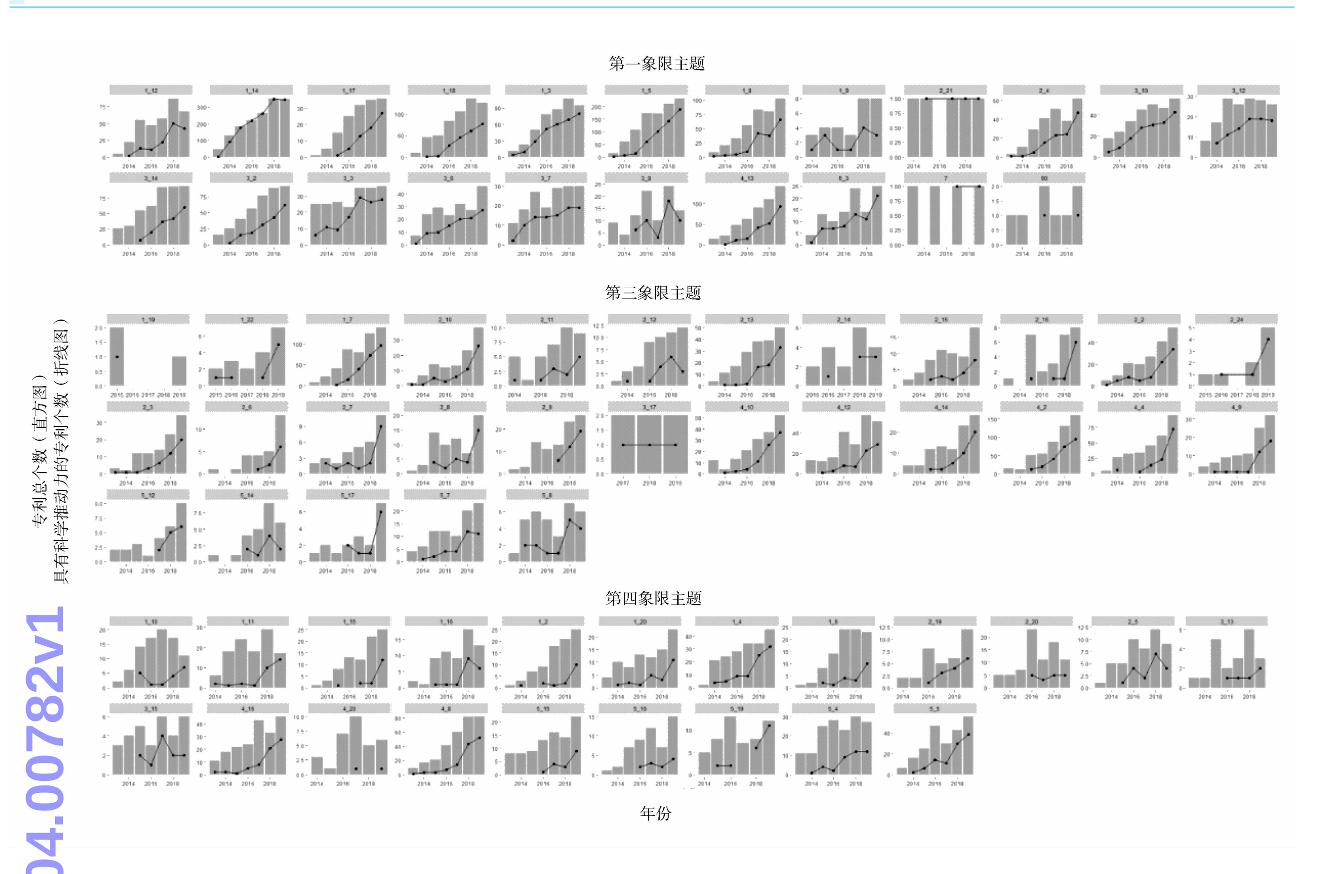


图 7 第一、三、四象限主题 2013 – 2019 年专利数量走势

算法使论文 – 专利关联网络形成科技主题聚类，构建相对科学推动力、相对技术成果数量指标，识别出科学推动力相对较大、而当前技术成果相对较少的科技主题作为技术机会。技术开发人员可关注这些技术机会，着重从科学研究中汲取理论知识与相关理念，应用到相应的技术创新中去。

在 3D 打印的案例研究中，该方法有效地形成了论文、专利的文档向量表示，构建了论文 – 专利关联网络与科技主题网络，最终识别了科学推动下的技术机会，并验证了所识别的机会具有一定的技术潜力，证明了方法的有效性。不同于以往基于论文与专利关系识别技术机会的研究，该方法无需将科学主题与技术主题人工配对，在科技主题形成之前就将论文与专利联系起来，减少了对主观判断的部分依赖，使过程更加自动化，也使技术颗粒更加精细化。

本研究也存在一些不足之处：①人为设定余弦相似度阈值是来筛选论文 – 专利关联网络，可能会把一些接近阈值但确有关联的关系排除，造成网络中连边的损失。在未来的研究中，会设计更完善的方法来构建关联网络。②科学与技术的关系是相互的，科学对技术具有推动力，技术对科学也具有拉动作用，本文只考虑了前者，在未来的研究中，可以考虑技术对科学的拉动作用，挖掘科研机会。③若在基于本文的网络构建方法上，加入引证关系，会使网络更加精确。但在实际尝试中，由于专利对论文的引用较少，导致网络十分稀疏，对文本网络的影响微乎其微，因此目前只使用文本的方法构建网络。未来会在研究中尝试解决这个问题，采用合适的方法将引证关系引入基于文本的网络中，使网络更加优化。④在创新链条中，除了科学研究与技术开发两个环节，还有商业应用环节，三者之间交互影响，未来可把技术机会分析扩展到创新全链条，结合多个环节更全面地识别技术机会。

参考文献：

[1] SCHUMPETER J A. The theory of economic development[J]. Journal of political economy, 1911, 1(2):170 – 172.

[2] PENROSE E. The Theory of the Growth of the Firm[M]. New Jersey:Blackwell, 2009.

[3] PORTER A L, MICHAEL J, DETAMPE L. Technology opportunities analysis[J]. Technological forecasting & social change, 1995, 49(3):237 – 255.

[4] PORTER A L, JIN X Y, GILMOUR J E, et al. Technology opportunities analysis - integrating technology monitoring, forecasting, and assessment with strategic-planning. [J]. SRA-Journal of the society of research, 1994, 26(2):21 – 31.

[5] ROBINSON D K R, HUANG L, GUO Y, et al. Forecasting Innovation Pathways (FIP) for new and emerging science and technologies[J]. Technological forecasting and social change, 2013, 80 (2):267-285.
[6] SHIBATA N, KAJIKAWA Y, SAKATA I. Extracting the commercialization gap between science and technology - Case study of a solar cell[J]. Technological forecasting and social change, 2010, 77 (7):1147-1155.
[7] DING Y, CHOWDHURY G G, FOO S. Journal as markers of intellectual space: journal co-citation analysis of Information Retrieval area, 1987-1997[J]. Scientometrics, 2000, 47(1):55-73.
[8] SHIBATA N, KAJIKAWA Y, SAKATA I. Detecting potential technological fronts by comparing scientific papers and patents[J]. Foresight, 2011, 13(5):51-60.
[9] TAKANO Y, KAJIKAWA Y. Extracting commercialization opportunities of the Internet of Things: measuring text similarity between papers and patents [J]. Technological forecasting and social change, 2019, 138(1):45-68.
[10] LI X, XIE Q, DAIM T, et al. Forecasting technology trends using text mining of the gaps between science and technology: the case of perovskite solar cell technology[J]. Technological forecasting and social change, 2019, 146(9):432-449.
[11] WANG M Y, FANG S C, CHANG Y H. Exploring technological opportunities by mining the gaps between science and technology: microalgal biofuels [J]. Technological forecasting and social change, 2015, 92(3):182-195.
[12] 王京安, 汤月, 王坤. 基于 Citespace 的技术机会发现研究——以物联网技术发展为例[J]. 现代情报, 2018, 38(2):130-137,170.
[13] 黄鲁成, 王静静, 李欣, 等. 基于论文和专利的钙钛矿太阳能电池的技术机会分析[J]. 情报学报, 2016, 35(7):686-695.
[14] 王坤, 王京安, 汤月, 等. 基于专利和科技论文的技术机会识别研究——以金属 3D 打印技术为例[J]. 科技管理研究, 2018, 38(7):73-79.
[15] PRICE M. Recent Studies in the Restoration and Eighteenth Century[J]. Studies in english literature, 1965, 5(3): 553-574.
[16] NARIN F, HAMILTON K S, OLIVASTRO D. The increasing linkage between U. S. technology and public science[J]. Research policy, 1997, 26(3):317-330.
[17] NARIN F, OLIVASTRO D. Status report: Linkage between technology and science [J]. Research policy, 1992, 21 (3): 237-249.
[18] BRESCHI S, CATALINI C. Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks[J]. Research policy, 2010, 39(1):14-26.
[19] Rosenberg N. Inside the Black Box[M]. Cambridge: Cambridge University Press,2010.
[20] KLEVORICK A K, LEVIN R C, NELSON R R, et al. On the sources and significance of interindustry differences in technological opportunities[J]. Research policy, 1995, 24(2):185-205.
[21] MANSFIELD E. Academic research and industrial innovation[J]. Research policy, 1991, 20(1):1-12.
[22] MEYER M. Tracing knowledge flows in innovation systems-an informetric perspective on future research science-based innovation [J]. Economic systems research, 2002, 14(4):323-344.
[23] PETRESCU A S. Science and technology for economic growth, new insights from when the data contradicts desktop models 1[J]. Review of policy research, 2009, 26(6):839-880.
[24] 赖院根. 期刊论文与专利文献的链接研究[J]. 图书情报知识, 2011(1):63-69.
[25] MCMILLAN G S, NARIN F, DEEDS D L. An analysis of the critical role of public science in innovation: the case of biotechnology [J]. Research policy, 2000, 29(1):1-8.
[26] NARIN F, NOMA E. Is technology becoming science? [J]. Scientometrics, 1985, 7(3):369-381.
[27] TIJSSEN, ROBERT JW. Global and domestic utilization of industrial relevant science: patent citation analysis of science-technology interactions and knowledge flows[J]. Research policy, 2001, 30 (1):35-54.
[28] 赵黎明, 高杨, 韩宇. 专利引文分析在知识转移机制研究中的应用[J]. 科学学研究, 2002, 20(3):297-300.
[29] 刘立, 王耀德. 从专利引文看公共科学对技术创新的重要作用[J]. 科学学研究, 2003, 21(4):428-432.
[30] 殷媛媛. 基于论文专利引证关系的科学技术互动研究——以立体显示为实证分析[J]. 图书情报工作, 2012, 56(16):65-70,74.
[31] FLEMING L, SORENSON O. Science as a map in technological search[J]. Strategic management journal, 2004, 25(89):909-928.
[32] HELLMANN T. The role of patents for bridging the science to market gap[J]. Journal of economic behavior & organization, 2007, 63(4):624-647.
[33] 任智军, 乔晓东, 徐硕, 等. 基于数据挖掘的技术机会发现模型研究[J]. 情报杂志, 2015, 34(6):174-177,190.
[34] 张福俊, 叶权慧, 于路云. 基于知识图谱的海洋科学领域技术机会分析[J]. 科技管理研究, 2017, 37(24):165-170.
[35] 王兴旺, 董珏, 余婷婷, 等. 基于多种类型信息计量分析的前沿技术预测方法研究[J]. 情报杂志, 2018, 37(10):70-75,89.
[36] 韩燕, 彭爱东. 基于技术形成三要素的技术机会识别研究——以医疗服务机器人领域技术为例[J]. 情报理论与实践, 2020, 43(5): 156-162.
[37] LE Q, MIKOLOV T. Distributed representations of sentences and documents[C]//ICML. Proceedings of the 31st International Con-

ference on Machine Learning. Beijing:JMLR,2014.

[38] BLONDEL V D, GUILLAUME J-L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics: theory and experiment, 2008(10):P10008.

[39] CARPENTER M P, COOPER M, NARIN F. Linkage Between Basic Research Literature and Patents[J]. Research management, 1980, 23(2):30–35.

[40] HUANG Y, ZHU D, QIAN Y, et al. A hybrid method to trace technology evolution pathways: a case study of 3D printing[J]. Scientometrics, 2017, 111(1):185–204.

[41] 王金风, 吴敏, 岳俊举, 等. 创新过程的技术机会识别路径研究——基于专利挖掘和形态分析[J]. 情报理论与实践, 2017, 40(8):82–86.

作者贡献说明:

韩晓彤:方法设计、程序编写、数据处理、论文撰写;

朱东华:论文思路指导;

汪雪锋:论文思路指导、方法设计、写作指导、论文修改。

Research on the Method of Technology Opportunity Discovery Promoted by Science

Han Xiaotong Zhu Donghua Wang Xuefeng

School of Management and Economics, Beijing Institute of Technology, Beijing 100081

Abstract: [**Purpose/Significance**] The close relationship between science and technology makes it more reasonable and efficient to analyze technology opportunities by combining papers and patents rather than using single data. This paper makes the generation of science-technology relationship more automatic, reduces the dependence on subjective judgment, and makes the technology units smaller. The purpose is to provide R&D suggestions for technology researchers, and help apply theories and ideas from scientific research to technological innovation. [**Method/Process**] The abstract texts of papers and patents were represented by Doc2vec vector, which were associated into a network through text similarity, and then science-technology clusters were generated based on Louvain algorithm to identify technology opportunities promoted by scientific research. Finally, 3D printing technology was taken as a case for empirical research. [**Result/Conclusion**] Several technology opportunities promoted by scientific research are identified, and it is verified that the identified opportunities have technological potential, which proves the feasibility and effectiveness of the method.

Keywords: technology opportunity analysis scientific impetus patent analysis Doc2vec Louvain algorithm